Dual Robust Unbiased Multi-View Clustering for Incomplete and Unpaired Information

Liang Zhao[1] , **Ziyue Wang**[1] , **Chuanye He**[1] , **Qingchen Zhang**[*, 2] and **Bo Xu**[1]
[1]School of Software, Dalian University of Technology, Dalian, Liaoning, China
[2]School of Computer Science and Technology, Hainan University, Haikou, Hainan, China
{liangzhao, BoXu}@dlut.edu.cn, {wangziyue6306, 2838527538}@mail.dlut.edu.cn,
zhangqingchen@hainanu.edu.cn

Abstract

Recently, multi-view data has gradually attracted attention. However, real-world applications often face Partial View-aligned Problem (PVP) and Partially Sample-missing Problem (PSP) due to data loss or corruption. Existing methods addressing PVP typically focus only on learning from the information of aligned data, while ignoring unaligned data where samples exist but lack alignment relationships. This introduces PSP, which does not inherently exist in the data, leading to biased learning of the data's information. For PSP, due to varying degrees of missing data, incomplete spatial structures can cause clustering centers-shifted problem, resulting in the model learning incorrect correspondences and biased spatial structures.To tackle them, we propose a novel method called Dual Robust Unbiased Multi-View Clustering for Incomplete and Unpaired Information (DRUMVC). To our knowledge, this is the first noise-robust and unbiased multi-view clustering method capable of simultaneously addressing both PVP and PSP. Specifically, DRUMVC leverages aligned and complete samples as a bridge to construct high-quality correspondences for samples lacking cross-view relationship information due to PVP or PSP. Additionally, we employ a dual noise-robust contrastive learning loss to mitigate the impact of noise potentially introduced during the pair construction. Experiments on several challenging datasets demonstrate the superiority of our proposed method.

1 Introduction

Multi-view clustering (MVC) aims to learn more accurate common representations of multi-view data by exploiting both the consistency and complementarity of multi-view information. However, the above methods heavily rely on the assumptions of view consistency that is, the correspondences between multi-view data for the same object are complete and instance completeness that is, the multi-view data itself is complete without missing. In real-world scenarios, however, the correspondences between multi-view data are often incomplete. PVC [Huang *et al.*, 2020] provides a classic example: due to street surveillance cameras being distributed at different locations along a street, the target of interest may be captured at different times and locations by different cameras. Due to such spatial or temporal complexities, it is difficult to obtain cross-view correspondences, leading to the issue of PVP. Similarly, due to issues like unstable data transmission or storage device failures, data itself may be lost, resulting in the issue of PSP. When multi-view data is affected by PVP or PSP, existing MVC methods tend to produce suboptimal results.

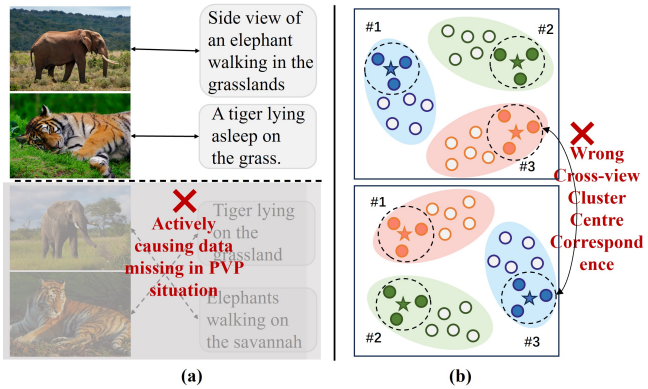

Figure 1: Ignoring unaligned data is equivalent to actively causing data to be missing in (a). Cluster centres learned from incomplete multiview data may be shifted and lead to incorrect correspondences in (b).

For PVP, most existing studies rely on contrastive learning to reconstruct the alignment relationships of unaligned data by learning the alignment relationships of known aligned data. However, unaligned data only lacks alignment relationships, while the data itself still exists. Therefore, focusing solely on the aligned data while ignoring the information contained in the unaligned data is unreasonable. Such an approach may inadvertently introduce PSP issues into multi-view data that originally only had PVP issues, as illustrated in Figure 1(a).

In recent years, various studies have proposed different methods to address PSP, such as matrix factorization-based methods[Li *et al.*, 2014; Zhao *et al.*, 2016; Shao *et*

al., 2015; Hu and Chen, 2019a], incomplete multi-kernel learning[Bach and Jordan, 2002; Liu *et al.*, 2020], and graph-based approaches[Wang *et al.*, 2019; Liu *et al.*, 2019]. However, due to the incompleteness of multi-view data, the distributions learned from incoherent data tend to be biased. Specifically, because of missing data, the cluster centers or prototypes of different classes within each view may shift. When extended to the multi-view, this results in incorrect correspondences between the cluster centers or prototypes of the same class across different views, as illustrated in Figure 1(b).

Moreover, most current methods addressing PSP or PVP, such as Deep Incomplete Multi-View Clustering (DIMVC) [Lin *et al.*, 2021a; Zhu *et al.*, 2019; Wang *et al.*, 2020; Wen *et al.*, 2021; Zhao *et al.*, 2024a; Zhao *et al.*, 2025] and Deep Partial View-Alignment Clustering (DPVC) [Yang *et al.*, 2021a; Wang *et al.*, 2024a; Zhao *et al.*, 2024b], adopt a contrastive learning approach. They treat the representations of different views of the same sample as positive pairs, while treating the remaining samples across different views as negative pairs. However, among these negative pairs, there may exist sample pairs that, while not belonging to the same sample, belong to the same class. For clustering tasks, the model needs to learn more about cluster-level separability. So from the perspective of clustering tasks and clusters, this negative pair construction strategy may introduce noisy information, i.e., false negatives, thereby leading to suboptimal clustering performance.

To address the aforementioned challenges, we propose a novel method called Dual Robust Unbiased Multi-View Clustering for Incomplete and Unpaired Information (DRUMVC), which simultaneously addresses PSP and PVP within a unified framework. Specifically, to tackle the issue of incorrect PSP introduction in PVP and the clustering center shift caused by missing data in PSP, we leverage complete and aligned data within PSP or PVP, rather than relying on clustering centers, as an anchor set to explore similarity relationships with the complete dataset. This approach avoids the learning of erroneous correspondence information caused by clustering center shifts.

Furthermore, we propagate anchor graph information across views based on the assumption of semantic consistency [Wang *et al.*, 2022a], thereby constructing high-quality cross-view correspondences for incomplete and unpaired data. This enables the integration of incomplete and unpaired data into the model's learning process, while also addressing clustering center shift issues by compensating for the missing cross-view correspondences in incomplete data.

Moreover, to mitigate the impact of noise in sample construction, we also design a dual noise-robust contrastive learning loss. This loss aims to reduce the introduction of noise and alleviate the effect of false negatives on model training, both during sample pair construction and throughout the contrastive learning process. Leveraging these modules, DRUMVC is capable of learning more complete and robust consistency information for multi-view data, thereby achieving class-level cross-view correspondence reconstruction. The contributions of this work can be summarized as follows:

- We propose an unbiased multi-view representation learning module that constructs an anchor graph for the entire dataset by leveraging aligned and complete data as the anchor set. Then through cross-view propagation, it establishes high-quality cross-view correspondences for unaligned and incomplete data, thereby incorporating data with PVP and PSP into the model's learning process. This approach effectively avoids issues such as the erroneous introduction of PSP in PVP methods and clustering center shifts.
- From the perspectives of sample pair construction and model learning, we design a dual noise-robust contrastive learning loss. By exploiting the neighborhood information of complete and aligned data, the method filters out most of the noise introduced by random negative sample selection. Additionally, based on the similarity assumption for positive and negative samples (i.e., samples of the same class should be close in high-dimensional space, while samples of different classes should be far apart), we construct a novel contrastive loss to mitigate or even eliminate the impact of false negatives during model training.
- We conduct extensive experiments on several commonly used real-world datasets under various scenarios, including PVP, PSP, and their simultaneous presence. The proposed method consistently achieves state-of-the-art performance. Moreover, the model demonstrates exceptional clustering performance even when the data is severely corrupted.

2 Method

In this section, we provide a detailed introduction to the proposed DRUMVC, designed to simultaneously address both incompleteness and unalignment issues. First, we describe PSP and PVP that needs to be addressed. Then, the implementation process of the model is elaborated in the subsequent three sections. The overall model architecture is illustrated in Figure 2.

2.1 Problem Formulation

For $\{\mathrm{X}^{(v)}\}_{v=1}^{V} = \{\mathrm{x}_1^{(v)}, \mathrm{x}_2^{(v)}, \ldots, \mathrm{x}_N^{(v)}\}_{v=1}^{V}$ a given multi-view dataset, where V represents the number of views, N denotes the number of samples in each view, and $\mathrm{X}^{(v)} \in \mathbb{R}^{N \times D}$.Due to the presence of PVP and PSP, the original data can be divided into $\{S^{(v)}\}_{v=1}^{V} = \{s_1^{(v)}, s_2^{(v)}, \ldots, s_{N_{AC}}^{(v)}\}_{v=1}^{V}$ and $\{\mathrm{W}^{(v)}\}_{v=1}^{V} = \{\mathrm{w}_1^{(v)}, \mathrm{w}_2^{(v)}, \ldots, \mathrm{w}_{N_{UI}}^{(v)}\}_{v=1}^{V}$, where $N_{AC} + N_{UI} = N$. The sets $\{S^{(v)}\}_{v=1}^{V}$ and $\{W^{(v)}\}_{v=1}^{V}$ represent complete and aligned data, and data affected by PSP, PVP, or both, respectively. In this paper, we reorganize the multi-view dataset such that the complete and aligned data is placed in the first part, while the data with PVP, PSP, or both is placed in the latter part. Thus, we have $\{X^{(v)}\}_{v=1}^{V} = \{S^{(v)}, W^{(v)}\}_{v=1}^{V}$.

2.2 Mutual Information-like Representation Learning

Data samples from different views often differ in dimensionality due to variations in the inherent structure of the

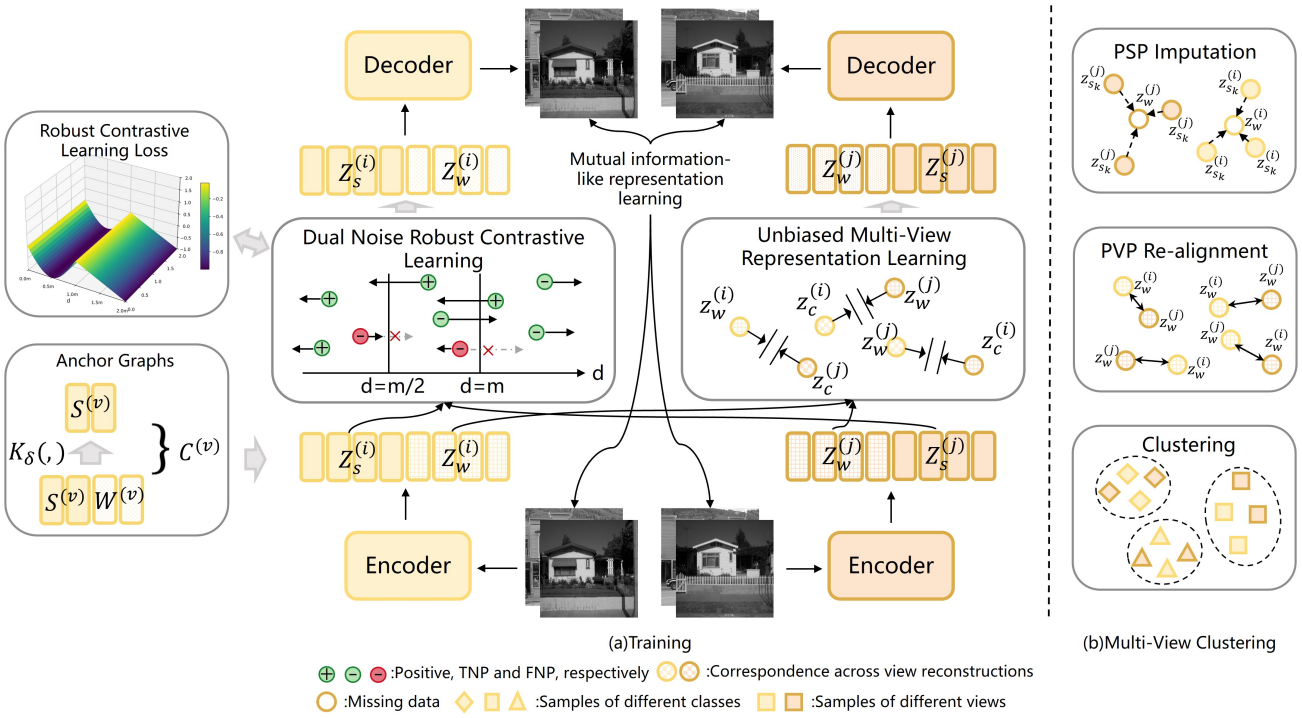

Figure 2: The overall framework of the proposed DRUMVC. We utilize the combined effect of three modules: Mutual Information-like Representation Learning, Unbiased Multi-View Representation Learning, and Dual Noise Robust Contrastive Learning to learn accurate and complete latent representations in the presence of both PSP and PVP. After training, we impute and realign latent representations and apply them to the clustering task.

data. Additionally, real-world multi-view data typically exhibit high dimensionality. As a result, directly processing the raw data not only suffers from the impact of redundant information but also incurs high computational complexity in terms of both time and space. To address these, We will derive the latent representation $Z^{(v)}$ for each view using a set of view-specific encoders:

$$Z^{(v)} = f_{\theta_f}^{v}\left(X^{(v)}\right) \tag{1}$$

where θ_v is the network layer parameters of encoder.

Since the original data $\{\mathrm{X}^{(v)}\}_{v=1}^{V}$ can be divided into two parts, $\{\mathrm{S}^{(v)}\}_{v=1}^{V}$ and $\{\mathrm{W}^{(v)}\}_{v=1}^{V}$, the corresponding $\{Z^{(v)}\}_{v=1}^{V}$ can similarly be divided into two parts, $\{Z_s^{(v)}\}_{v=1}^{V}$ and $\{Z_w^{(v)}\}_{v=1}^{V}$. It is worth noting that, in recent years, reconstruction loss based on information theory [Oord *et al.*, 2018; He *et al.*, 2020] has garnered increasing attention:

$$\begin{aligned} L_{it} = &-\sum_{t=1}^{N_{AC}}\left(I\left(Z_{s_t}^{(i)}, Z_{s_t}^{(j)}\right)\right) \\ &-\alpha\left(H\left(Z_{s_t}^{(i)}\right)+H\left(Z_{s_t}^{(j)}\right)\right) \end{aligned} \tag{2}$$

where I represents mutual information, and N_{AC} refers to the number of aligned and complete samples. By maximizing the mutual information between data from different views, the consistency information across views can be effectively explored.

Leveraging the concept of mutual information, we no longer directly feed the latent representations of multiple views into their respective decoders. Instead, we concatenate the latent representations of multiple views, denoted as $\{Z_s^{(v)}\}_{v=1}^{V}$, and feed the concatenated representation into the decoders. Consequently, the overall loss for this module is:

$$\begin{aligned} L_{MIL} &= \frac{1}{V}\sum_{v=1}^{V} L_{MIL}^{(v)} \\ &= \frac{1}{V}\sum_{v=1}^{V}\|S^{(v)} - g_{\theta_g}^{v}([Z_s^{(1)}, Z_s^{(2)}, \cdots, Z_s^{(v)}])\|_F^2 \end{aligned} \tag{3}$$

where $[...]$ represents the concatenation operation for multi-view data, and θ_g denotes the learnable parameters of the decoder.

2.3 Unbiased Multi-View Representation Learning

When multi-view data simultaneously suffers from both incompleteness and unalignment, the situation becomes even more complex. The corrupted data is difficult to explore for multi-view consistency information because it not only has missing data but also lacks cross-view correspondences. Furthermore, when the proportion of incomplete and unaligned data exceeds 50%, it becomes unreasonable to infer overall data information using only a small portion of the data. Therefore, we propose an unbiased multi-view representation learning approach to complete the missing and misaligned

data, enabling the model to go beyond learning only from the aligned and complete portions of the data.

We use complete and aligned data $\{S^{(v)}\}_{v=1}^{V}$ as the anchor set $\{\mathrm{A}^{(v)}\}_{v=1}^{V}$ to measure the similarity relationship between the data within each view and the anchors. The anchor graph $C^{(v)} \in R^{N \times N_{AC}}$ is defined as:

$$\mathrm{C}_{ij}^{(v)} = \begin{cases} \frac{K_\delta(x_i^{(v)}, \mathrm{a}_j^{(v)})}{\sum_{j \in \langle i \rangle^v} K_\delta(x_i^{(v)}, \mathrm{a}_j^{(v)})}, & \forall j \in \langle i \rangle^v \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $\langle i \rangle^v$ denotes the index set of the m ($m << n$) nearest anchors to $x_i^{(v)}$, determined using a distance function such as the l_2 distance. Although the anchor graph $C^{(v)}$ has N_{AC} columns, each row contains only m non-zero values that sum to 1. To conserve space, the zero values can be omitted, reducing the graph to the form $R^{N \times m}$. However, for the sake of clarity and to facilitate understanding in subsequent descriptions, we choose to retain the zero values in the paper. Furthermore, $K_\delta\left(x_i^{(v)}, a_j^{(v)}\right) = exp(-D^2(x_i^{(v)}, a_j^{(v)})/\sigma^2)$ is a commonly used Gaussian kernel function with a bandwidth parameter δ.

We achieve learning from incomplete and unaligned data by leveraging the information contained in the anchor graph $C^{(v)}$. Specifically, As AMCP [Zhao *et al.*, 2024c] says, if the corresponding samples in different views of multi-view data describe the same object in different forms, these corresponding samples across multiple views are not only similar themselves but also exhibit similar neighborhood information. When extended to incomplete and unaligned data, even if a sample in one view cannot find its corresponding sample due to PVP or PSP, we can reasonably infer the spatial neighborhood information of its corresponding sample based on the neighborhood information of the given sample. This enables the cross-view transfer of neighborhood structures.

Moreover, due to our anchor selection strategy and the consistency assumption of multi-view data [Luo *et al.*, 2018], aligned and complete data across different views are expected to be similar. Consequently, by leveraging the alignment relationships of anchors and the cross-view transfer of neighborhood structures, we can construct high-quality cross-view correspondences for misaligned and incomplete data.

$$Z_c^{(j)} = C_{[N_{AC}:,:]}^{(i)} \cdot Z_s^{(j)} \tag{5}$$

where $[N_{AC} :, :]$ refers to the rows of anchor graph information corresponding to incomplete and unaligned data in the i-th view. $Z_s^{(j)}$ represents the aligned and complete samples and anchor set in the j-th view. Meanwhile, $Z_c^{(j)}$ is constructed in the j-th view based on the neighborhood information of incomplete and unaligned data from the i-th view, forming cross-view correspondences.

Thus, incomplete and unaligned data can also be incorporated into the scope of model learning. However, as cross-view correspondences are constructed through the transfer of cross-view neighborhood structures, minimizing the spatial distance between them is not appropriate. Instead, we propose optimizing it to a boundary using a bounded contrastive learning loss.

$$L_t^{(v)} = \frac{1}{V-1} \sum_{j \neq i}^{V} |\left\| z_{w_t}^{(i)} - z_{c_t}^{(j)} \right\|_2 - a| \tag{6}$$

where a is the boundary of similarity. Extending this to the entire view, the overall loss of this module is:

$$L_{UMR} = \frac{1}{VN_{UI}} \sum_{v=1}^{V} \sum_{t=1}^{N_{UI}} L_t^{(v)} \tag{7}$$

2.4 Dual Noise Robust Contrastive Learning

To mitigate or even eliminate the impact of false negatives, we designed a Dual Noise-Robust Contrastive Learning (DRC) loss from two perspectives: sample pair construction and the contrastive learning process.

$$L_{DRC} = \frac{1}{N_{AC}V} \sum_{v=1}^{V} \sum_{i=1}^{N_{AC}} (PL_{p_i}^{(v)} + (1-P)L_{n_i}^{(v)}) \tag{8}$$

where $P = 1/0$ for positive/negative pairs, $L_{p_i}^{(v)}$ will take effects when the pairs are positive, and $L_{n_i}^{(v)}$ acts on negative pairs.

For the anchor graph $C^{(v)}$ computed in the above module, since aligned and complete data are used as the anchor set, the similarity relationships corresponding to this part of the data in $C^{(v)}$ can essentially be transformed into the spatial structural information of the aligned and complete data within the view. In other words, each sample can derive its similarity relationships with its top m closest samples through $C^{(v)}$.

To further process the aligned and complete samples in each view, we define an index set $\langle k \rangle_i^v$, to store the index information of the top K ($K \leq m$) nearest neighbors for each sample. Additionally, based on the assumption of semantic consistency [Wang *et al.*, 2022b], neighborhood information within a view can be transferred across views and converted into adjacency relationships with cross-view samples.

Thus, for aligned and complete data, besides the explicit cross-view corresponding samples that can serve as positive sample pairs, the cross-view neighborhood information also contains implicit positive sample pair information. Finally, we mine the consistency information of multi-view data by minimizing the distance between these relationships:

$$\begin{aligned} L_{p_i}^{(v)} = & \frac{1}{V-1} \sum_{j \neq i}^{V} \left\| z_{s_t}^{(i)} - z_{s_t}^{(j)} \right\|_2^2 \\ & + \beta \cdot \frac{1}{(V-1)K} \sum_{j \neq i}^{V} \sum_{k=1}^{K} \left\| z_{s_t}^{(i)} - z_{p_k}^{(j)} \right\|_2^2 \end{aligned} \tag{9}$$

where $z_{s_t}^{(i)}$ represents the t-th aligned and complete sample in the i-th view, while $z_{p_k}^{(j)}$ denotes the k-th indirect positive sample identified in the j-th view based on the cross-view index set $\langle k \rangle_i^v$. β is a hyperparameter.

Additionally, by learning positive sample pairs, the model can further leverage the consistency information across multiple views, improving the accuracy and reliability of cross-view information transfer in unbiased multi-view representation learning.

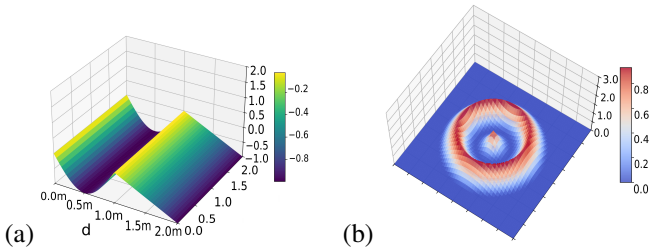

Figure 3: The performance surface of the loss function $L_{n_i}^{(v)}$ in 2D and 3D

Moreover, the cross-view index set $\langle k \rangle_i^v$ is not only utilized to mine indirect positive sample pair information across views but also serves to reduce the introduction of noisy information during sample pair construction. Specifically, for each aligned and complete sample $z_{s_t}^{(i)}$ as the anchor, M negative samples $z_n^{(j)}$ are randomly selected from the aligned and complete data across views, with the constraint that the cross-view negative samples $n \notin \langle k \rangle_t^i$. By preliminarily filtering neighborhood information, this approach effectively reduces the introduction of false negative pair information to some extent. Furthermore, we designed a noise-robust contrastive learning loss to mitigate or even eliminate the impact of false negative pairs during the model training process:

$$L_{n_i}^{(v)} = \begin{cases} \sin\left(\frac{\pi}{m_d}(2m_d - d)\right), 0 < d < m_d \\ m_d - d + c, \quad m_d \leq d \leq 2m_d \end{cases} \tag{10}$$

where d refers to the Euclidean distance between the negative sample pair $z_{s_t}^{(i)}$ and $z_n^{(j)}$, c is a constant and m_d is a distance margin that is calculated only once during the initial stage using the following formula:

$$m_d = \frac{1}{N_p}\Sigma \left\| z_{s_t}^{(i)} - z_{s_t}^{(j)} \right\|_2^2 + \frac{1}{N_n}\Sigma \left\| z_{s_t}^{(i)} - z_n^{(j)} \right\|_2^2 \tag{11}$$

where N_p refers to the total number of explicit positive sample pairs, while N_n represents the total number of negative sample pairs.

Next, we will demonstrate the noise robustness of the loss function using the visualizations in Figure 3. When the distance between negative sample pairs lies within the range $[0, m_d]$, the gradient of our loss function differs from that of conventional contrastive learning loss functions. The latter simply aims to increase the distance between negative pairs, failing to mitigate the impact of noisy negative pairs. For false negative pairs (FNP) within $[0, m_d/2]$, they are considered to belong to the same class but not the same sample, with relatively small distances. While the objective of reconstructing class-level relationships may require reducing these distances as much as possible, it is still necessary to preserve the discriminative information within the same class cluster to extract more accurate spatial structural information. Therefore, for FNP within $[0, m_d/2]$, we moderately increase their distance.For FNP within the range $[m_d/2, m_d]$, instead of simply pushing them apart as in conventional contrastive learning loss functions, we reverse their gradients and gradually pull them closer. This approach enhances the model's robustness to noise introduced by random sampling. For negative pairs within the range $[m_d, 2m_d]$, we directly push their distances apart.

Considering that a small number of TNP may exist within $[0, m_d]$, the loss functions designed for explicit and indirect positive pairs, along with the unbiased multi-view representation learning module, enable the model to learn sufficient consistency information between samples. For TNP with relatively large distances, the model has already learned the discriminative information between negative sample pairs, giving it the ability to distinguish true negative pairs. Additionally, compared to $L_{p_i}^{(v)}$, which focuses on the quadratic term of the sample pairs, $L_{n_i}^{(v)}$ focuses only on the linear term of the distance for negative pairs. Furthermore, the gradient of $L_{n_i}^{(v)}$ within $[0, m_d]$ has a maximum value of just 1. As a result, the small number of TNP within $[0, m_d]$ will not be optimized in the wrong direction due to the small reverse gradient.

Overall, the loss function is as follows:

$$L = L_{MIL} + \lambda L_{UMR} + L_{DRC} \tag{12}$$

Once the model converges, the cross-view correspondence for incomplete data is determined by calculating the distance matrix D between views. The mean of the neighboring data points is used for imputation. For PVP, the nearest sample across views is selected as the correspondence.

3 Experiments

3.1 Experimental Settings

In this study, we utilized four multi-view datasets: Scene15 [Fei-Fei and Perona, 2005], Reuters [Amini *et al.*, 2009], NoisyMNIST [Wang *et al.*, 2015], and MNIST-USPS [Peng *et al.*, 2019]. A brief description of each dataset is provided in Table 1.

In this paper, we conduct experiments on multi-view datasets with an alignment rate of 0.5, a completeness rate of 0.5, and various scenarios of unaligned incomplete multi-view datasets. The primary evaluation metrics include Accuracy (ACC), Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI). Higher values for these metrics indicate better clustering performance.

3.2 Comparisons with State of the Arts

Based on various settings, we compared DRUMVC with 18 latest state-of-the-art multi-view clustering baselines, including BMVC [Zhang *et al.*, 2018], AE2-Nets [Zhang *et al.*, 2019], PVC [Huang *et al.*, 2020], MvCLN [Yang *et al.*, 2021b], EGPVC [Zhao *et al.*, 2023], GCFAgg [Yan *et al.*, 2023], CMK [Liu *et al.*, 2023], SURE [Yang *et al.*, 2022],

Datasets	Samples	Classes	Features
Scene15	4485	15	{20,59}
Reuters	18758	6	{10,10}
NoisyMNIST	30000	10	{784,784}
MNISTUSPS	5000	10	{784,256}

Table 1: Statistics of the datasets.

Setting	Method	Scene15			Reuters			NoisyMNIST			MNIST-USPS		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Unpaired	BMVC	36.81	36.55	20.20	38.15	11.57	12.07	28.34	24.69	14.19	36.90	15.90	12.10
	AE2-Nets	28.56	26.58	12.96	35.49	10.61	8.07	38.25	34.32	22.02	37.60	23.90	16.10
	PVC	37.88	39.12	20.63	42.07	20.43	16.95	81.84	82.29	82.03	86.54	78.08	74.60
	MvCLN	38.53	39.90	24.26	50.16	30.65	24.90	91.05	84.15	83.56	89.96	81.36	80.40
	EGPVC	37.30	38.82	19.21	39.49	18.68	14.19	-	-	-	78.52	70.50	61.11
	GCFAgg	41.32	42.29	24.67	58.18	38.54	34.42	91.89	85.03	85.18	90.17	82.89	83.21
	CMK	40.51	42.12	25.38	59.48	38.94	31.54	93.60	86.41	86.89	92.35	83.72	84.38
	SURE	40.32	40.33	23.08	49.99	29.46	24.60	95.17	88.24	89.72	92.14	82.83	83.47
	OTGM	41.63	42.05	22.57	51.82	27.83	15.23	-	-	-	87.62	83.86	80.74
	NCL	44.57	**43.93**	26.46	64.51	**43.88**	38.87	96.00	89.93	91.46	95.60	89.23	90.50
	Ours	**45.33**	42.83	**26.79**	**65.13**	43.59	**39.70**	**96.33**	**90.61**	**92.15**	**96.06**	**90.46**	**91.49**
Incomplete	AE2-Nets	22.44	23.43	9.56	29.08	7.55	4.84	29.88	23.78	11.81	40.90	29.30	19.70
	PMVC	25.47	25.37	11.31	29.32	7.42	4.42	33.13	25.49	14.62	60.50	47.10	39.80
	UEAF	28.95	26.92	8.37	33.32	20.06	12.19	37.45	34.42	25.71	63.32	58.86	49.23
	DAIMC	27.00	23.47	10.62	40.94	18.66	15.04	33.81	26.42	15.96	55.20	49.60	38.60
	EERIMVC	31.50	31.11	14.82	29.77	12.01	4.21	55.62	45.92	36.76	65.20	55.70	48.90
	COMPLETER	39.50	42.35	23.51	34.61	17.53	2.93	80.01	75.23	70.66	88.91	89.52	85.31
	SURE	39.60	41.58	23.49	47.18	30.89	23.32	92.34	84.99	84.31	92.34	84.99	84.31
	DSIMVC	30.56	35.47	17.24	39.87	19.61	17.13	57.47	55.12	44.08	96.71	91.82	92.98
	ProImp	41.58	42.86	25.31	51.89	35.54	28.53	94.86	87.43	89.08	96.81	91.85	93.06
	DIVIDE	45.53	**45.53**	28.35	54.70	37.30	28.60	51.82	49.24	29.95	92.34	83.77	83.87
	Ours	**47.34**	44.72	**28.48**	**65.27**	**44.27**	**40.54**	**95.18**	**88.26**	**89.78**	**97.38**	**93.28**	**94.28**

Table 2: The clustering performance on four multi-view benchmarks with 0.5 align or 0.5 complete rate.

OTGM [Wang *et al.*, 2024b], NC3L [Qian *et al.*, 2024], PMVC [Li *et al.*, 2014], UEAF [Wen *et al.*, 2019], DAIMC [Hu and Chen, 2019b], EERIMVC [Liu *et al.*, 2020], COMPLETER [Lin *et al.*, 2021b], DSIMVC [Tang and Liu, 2022], ProImp [Li *et al.*, 2023], and DIVIDE [Lu *et al.*, 2024]. The best experimental results are in bold, with the second-best results marked using an underscore '_'. A dash '-' denotes impractical methods due to excessive time or memory consumption.

Since some of the comparison algorithms were not specifically designed for the PVP or PSP problems, we adopted the same approach as SURE. For partially aligned multi-view datasets, we used the Hungarian algorithm to realign the data before inputting it into the relevant models. In the case of incomplete multi-view datasets, we employed the mean value within each view for data imputation.

As shown in Table 2, our model achieves satisfactory experimental results in both settings compared to existing state-of-the-art algorithms for PVP or PSP. Although the performance improvement in unpaired setting over the NC3L model is not particularly significant, the lightweight nature of DRUMVC stands in contrast to the complex mathematical derivation process utilized by NC3L. Furthermore, in the challenging context of the incomplete multi-view dataset Reuters, DRUMVC exhibited a minimum of 19% improvement across all three metrics. These experimental findings validate the capability of DRUMVC to effectively mitigate data missing or unpaired challenges in multi-view datasets.

As of now, SURE is the only effective method that can address both PVP and PSP. Therefore, we follow up with experimental comparisons with it in a variety of PVP and PSP ratio scenarios. Table 3 reveals that, in comparison to SURE, DRUMVC yielded consistently optimal experimental outcomes across all datasets, accompanied by significant improvements. Notably, when data corruption was more severe, such as alignment rates of only 0.1 with completeness levels of 0.5 or 0.7, owing to its unbiased multi-view representation learning module (as detailed in Section 2.3), DRUMVC demonstrated an enhanced capacity to learn more comprehensive data information compared to other models. In particular, on the Noisymnist and NNIST-USPS datasets, when the alignment rate was 0.1 and the completeness level was 0.5, DRUMVC achieved a remarkable improvement of 81.15% in ACC, 92.30% in NMI, and an astonishing 156.67% in ARI on the Noisymnist dataset. Similarly, on the NNIST-USPS dataset, it achieves a 57.15% improvement in ACC, a 72.02% improvement in NMI, and a dramatic 110.66% improvement in ARI. These results not only demonstrate the effectiveness of DRUMVC in addressing the simultaneous occurrences of incompleteness and misalignment but also highlight our model's ability to maintain commendable clustering performance even in the presence of severe data degradation.

3.3 Ablation Study and Parameters Analysis

To validate the effectiveness of the designed modules, we conducted ablation experiments on three loss functions: L_{MIL}, L_{UMR}, and L_{DRC}. The L_{MIL} loss function leverages the concept of class mutual information to extract the consistency information of multi-view data; L_{UMR} is employed to ensure that the model learns unbiased spatial structural information and comprehensive consistency information; while L_{DRC} aims to reduce, or even eliminate, the impact of false negatives on noise. Table 4 indicates that the model's performance reaches its optimal level when all three losses are applied in conjunction.

Complete rate	Align rate	Method	Scene-15			Reuters			NoisyMNIST			MNIST-USPS		
			ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
0.5	0.1	SURE	25.22	19.32	10.06	36.89	15.43	11.18	38.47	23.89	17.42	42.90	25.41	19.60
		Ours	**30.79**	**21.07**	**12.88**	**47.31**	**18.81**	**19.25**	**69.69**	**45.94**	**44.71**	**67.42**	**43.71**	**41.29**
	0.3	SURE	33.62	26.82	15.45	40.44	22.04	17.92	73.65	53.57	52.22	72.67	50.96	49.69
		Ours	**35.05**	**27.00**	**16.00**	**54.58**	**25.18**	**24.77**	**78.93**	**58.03**	**58.84**	**78.78**	**58.20**	**58.54**
	0.5	SURE	36.57	30.43	17.84	43.98	20.65	16.65	84.94	67.54	69.42	83.02	65.30	66.07
		Ours	**40.58**	**32.08**	**20.45**	**58.58**	**30.51**	**29.96**	**85.77**	**68.82**	**70.93**	**85.88**	**69.42**	**71.09**
	0.7	SURE	37.30	35.05	21.71	44.74	23.45	19.83	90.60	77.72	80.26	90.18	77.33	79.37
		Ours	**41.05**	**35.73**	**21.81**	**59.68**	**32.85**	**32.04**	**91.21**	**78.86**	**81.44**	**91.22**	**79.47**	**81.72**
0.7	0.1	SURE	31.19	24.00	12.96	39.66	18.73	15.10	39.27	29.96	22.15	41.38	30.21	22.30
		Ours	**32.55**	**26.21**	**15.84**	**50.71**	**21.99**	**23.41**	**76.19**	**55.85**	**55.06**	**73.36**	**52.94**	**50.58**
	0.3	SURE	35.92	30.19	18.31	47.52	21.19	17.35	73.84	58.70	55.77	63.75	47.03	41.86
		Ours	**37.57**	**31.29**	**19.04**	**55.46**	**27.20**	**27.37**	**85.04**	**68.09**	**69.77**	**83.94**	**66.91**	**67.68**
	0.5	SURE	35.43	33.41	19.31	43.18	26.25	20.58	88.38	73.88	76.00	86.94	71.78	73.22
		Ours	**41.20**	**35.05**	**21.88**	**58.73**	**32.01**	**30.53**	**89.71**	**76.09**	**78.51**	**90.10**	**77.28**	**79.33**
	0.7	SURE	39.09	37.58	22.02	43.93	25.19	22.52	93.48	83.63	86.07	92.78	82.39	84.62
		Ours	**43.37**	**38.99**	**24.22**	**60.64**	**37.20**	**34.56**	**93.75**	**84.11**	**86.61**	**93.88**	**84.64**	**86.85**

Table 3: The clustering performance on four datasets under both PVP and PSP.

L_{MIL}	L_{UMR}	L_{DRC}	ACC	NMI	ARI
✓	✓		23.95	18.63	8.53
✓		✓	34.47	31.41	18.45
	✓	✓	36.72	31.64	18.78
✓	✓	✓	**40.58**	**32.08**	**20.45**

Table 4: Ablation studies on Scene15.

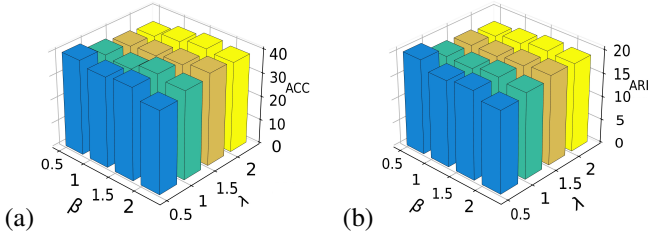

Figure 4: Analysis of model parameters on Scene-15.

Similarly, we evaluated the model's performance under various combinations of hyperparameters with 0.5 paired and complete rate. As shown in Figure 4, our model exhibits robust performance concerning the chosen hyperparameters.

3.4 Visualization Results

We present a visualization of the clustering results for SURE and our model DRUMVC on the Noisymnist datasets, both characterized by alignment and completeness rates of 0.5. Figure 5 shows that DRUMVC effectively learns more comprehensive spatial structural information and unbiased consistency information through its unbiased multi-view representation learning module. It also shows the dual noise-robust contrastive learning module significantly mitigates the impact of false negatives. Consequently, the clustering outcomes produced by DRUMVC do not exhibit instances of data partitioning within the same class, in contrast to SURE. Furthermore, due to the model's ability to learn complete data information, the cluster centroids within each class Cluster are more accurate.

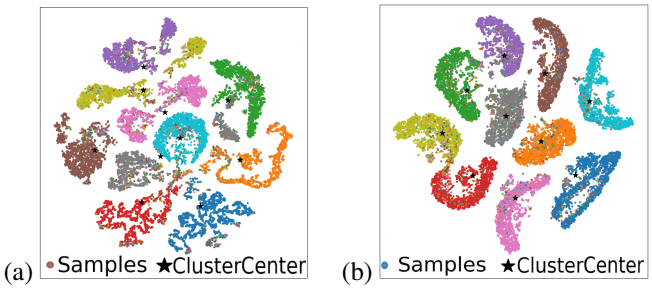

Figure 5: SURE as (a) and Ours as (b) Clustering Visualisation in NoisyMNIST.

4 Conclusion

To mitigate the erroneous introduction of PSP in PVP datasets, address the issue of cluster centroid displacement caused by data incompleteness, and tackle the presence of noise in sample pair construction, we presented a novel algorithm, DRUMVC, capable of simultaneously addressing both PSP and PVP challenges. By incorporating problematic data into the model's learning process, DRUMVC enables the model to acquire a more comprehensive understanding of the data information. Furthermore, by utilizing neighborhood information and spatial distance to implement a dual noise-robust contrastive loss, the algorithm effectively reduces the impact of noise on data information extraction. Through the synergistic interaction of multiple modules, DRUMVC achieves satisfactory clustering performance across various datasets.

Acknowledgments

This work is supported by the Science and Technol ogy Project of Liaoning Province (2024JH2/102600027, 2023JH2/101700363) and the Science and Technology Project of Dalian City (2024JJ12GX025, 2023JJ12SN029 and 2023JJ11CG005).

References

[Amini *et al.*, 2009] Massih R Amini, Nicolas Usunier, and Cyril Goutte. Learning from multiple partially observed views-an application to multilingual text categorization. *Advances in neural information processing systems*, 22, 2009.

[Bach and Jordan, 2002] Francis R Bach and M Jordan. Kernel independent component analysis: Journal of machine learning research. 2002.

[Fei-Fei and Perona, 2005] Li Fei-Fei and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 2, pages 524–531. IEEE, 2005.

[He *et al.*, 2020] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[Hu and Chen, 2019a] Menglei Hu and Songcan Chen. Doubly aligned incomplete multi-view clustering. *arXiv preprint arXiv:1903.02785*, 2019.

[Hu and Chen, 2019b] Menglei Hu and Songcan Chen. Doubly aligned incomplete multi-view clustering. *arXiv preprint arXiv:1903.02785*, 2019.

[Huang *et al.*, 2020] Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv, and Xi Peng. Partially view-aligned clustering. *Advances in Neural Information Processing Systems*, 33:2892–2902, 2020.

[Li *et al.*, 2014] Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28, 2014.

[Li *et al.*, 2023] Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. *arXiv preprint arXiv:2301.11045*, 2023.

[Lin *et al.*, 2021a] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11174–11183, 2021.

[Lin *et al.*, 2021b] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11174–11183, 2021.

[Liu *et al.*, 2019] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, En Zhu, Tongliang Liu, Marius Kloft, Dinggang Shen, Jianping Yin, and Wen Gao. Multiple kernel k k-means with incomplete kernels. *IEEE transactions on pattern analysis and machine intelligence*, 42(5):1191–1204, 2019.

[Liu *et al.*, 2020] Xinwang Liu, Miaomiao Li, Chang Tang, Jingyuan Xia, Jian Xiong, Li Liu, Marius Kloft, and En Zhu. Efficient and effective regularized incomplete multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence*, 43(8):2634–2646, 2020.

[Liu *et al.*, 2023] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Qing Liao, and Yuanqing Xia. Contrastive multi-view kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):9552–9566, 2023.

[Lu *et al.*, 2024] Yiding Lu, Yijie Lin, Mouxing Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled contrastive multi-view clustering with high-order random walks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 14193–14201, 2024.

[Luo *et al.*, 2018] Shirui Luo, Changqing Zhang, Wei Zhang, and Xiaochun Cao. Consistent and specific multi-view subspace clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[Oord *et al.*, 2018] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[Peng *et al.*, 2019] Xi Peng, Zhenyu Huang, Jiancheng Lv, Hongyuan Zhu, and Joey Tianyi Zhou. Comic: Multi-view clustering without parameter selection. In *International conference on machine learning*, pages 5092–5101. PMLR, 2019.

[Qian *et al.*, 2024] Shengsheng Qian, Dizhan Xue, Jun Hu, Huaiwen Zhang, and Changsheng Xu. Nonparametric clustering-guided cross-view contrastive learning for partially view-aligned representation learning. *IEEE Transactions on Image Processing*, 2024.

[Shao *et al.*, 2015] Weixiang Shao, Lifang He, and Philip S Yu. Multiple incomplete views clustering via weighted nonnegative matrix factorization with regularization. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 318–334. Springer, 2015.

[Tang and Liu, 2022] Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International Conference on Machine Learning*, pages 21090–21110. PMLR, 2022.

[Wang *et al.*, 2015] Weiran Wang, Raman Arora, Karen Livescu, and Jeff Bilmes. On deep multi-view representation learning. In *International conference on machine learning*, pages 1083–1092. PMLR, 2015.

[Wang *et al.*, 2019] Hao Wang, Linlin Zong, Bing Liu, Yan Yang, and Wei Zhou. Spectral perturbation meets incomplete multi-view data. *arXiv preprint arXiv:1906.00098*, 2019.

[Wang *et al.*, 2020] Qianqian Wang, Huanhuan Lian, Gan Sun, Quanxue Gao, and Licheng Jiao. icmsc: Incomplete cross-modal subspace clustering. *IEEE Transactions on Image Processing*, 30:305–317, 2020.

[Wang *et al.*, 2022a] Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen, and Yao Zhao. Graph contrastive partial multi-view clustering. *IEEE Transactions on Multimedia*, 25:6551–6562, 2022.

[Wang *et al.*, 2022b] Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen, and Yao Zhao. Incomplete multiview clustering via cross-view relation transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):367–378, 2022.

[Wang *et al.*, 2024a] Xibiao Wang, Hang Gao, Xindian Wei, Liang Peng, Rui Li, Cheng Liu, Si Wu, and Hau-San Wong. Contrastive graph distribution alignment for partially view-aligned clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 5240–5249, 2024.

[Wang *et al.*, 2024b] Xibiao Wang, Hang Gao, Xindian Wei, Liang Peng, Rui Li, Cheng Liu, Si Wu, and Hau-San Wong. Contrastive graph distribution alignment for partially view-aligned clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 5240–5249, 2024.

[Wen *et al.*, 2019] Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Hong Liu. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5393–5400, 2019.

[Wen *et al.*, 2021] Jie Wen, Zhihao Wu, Zheng Zhang, Lunke Fei, Bob Zhang, and Yong Xu. Structural deep incomplete multi-view clustering network. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 3538–3542, 2021.

[Yan *et al.*, 2023] Weiqing Yan, Yuanyang Zhang, Chenlei Lv, Chang Tang, Guanghui Yue, Liang Liao, and Weisi Lin. Gcfagg: Global and cross-view feature aggregation for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19863–19872, 2023.

[Yang *et al.*, 2021a] Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu, and Xi Peng. Partially view-aligned representation learning with noise-robust contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1134–1143, 2021.

[Yang *et al.*, 2021b] Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu, and Xi Peng. Partially view-aligned representation learning with noise-robust contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1134–1143, 2021.

[Yang *et al.*, 2022] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2022.

[Zhang *et al.*, 2018] Zheng Zhang, Li Liu, Fumin Shen, Heng Tao Shen, and Ling Shao. Binary multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1774–1782, 2018.

[Zhang *et al.*, 2019] Changqing Zhang, Yeqing Liu, and Huazhu Fu. Ae2-nets: Autoencoder in autoencoder networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2577–2585, 2019.

[Zhao *et al.*, 2016] Handong Zhao, Hongfu Liu, and Yun Fu. Incomplete multi-modal visual data grouping. In *IJCAI*, pages 2392–2398, 2016.

[Zhao *et al.*, 2023] Liang Zhao, Qiongjie Xie, Sontao Wu, and Shubin Ma. An end-to-end framework for partial view-aligned clustering with graph structure. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[Zhao *et al.*, 2024a] Liang Zhao, Xiao Wang, Zhenjiao Liu, Ziyue Wang, and Zhikui Chen. Learnable graph guided deep multi-view representation learning via information bottleneck. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[Zhao *et al.*, 2024b] Liang Zhao, Qiongjie Xie, Zhengtao Li, Songtao Wu, and Yi Yang. Dynamic graph guided progressive partial view-aligned clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Zhao *et al.*, 2024c] Liang Zhao, Yukun Yuan, Qiongjie Xie, and Ziyue Wang. Anchor based multi-view clustering for partially view-aligned data. In *2024 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–5. IEEE, 2024.

[Zhao *et al.*, 2025] Liang Zhao, Ziyue Wang, Xiao Wang, Zhikui Chen, and Bo Xu. Incomplete and unpaired multi-view graph clustering with cross-view feature fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 22786–22794, 2025.

[Zhu *et al.*, 2019] Pengfei Zhu, Xinjie Yao, Yu Wang, Binyuan Hui, Dawei Du, and Qinghua Hu. Multi-view deep subspace clustering networks. *arXiv preprint arXiv:1908.01978*, 2019.